

# Appendix A

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2005, 09:51:41 ; Search time 3878.11 Seconds  
(without alignments)  
9982.023 Million cell updates/sec

Title: US-10-511-270-3  
Perfect score: 1017  
Sequence: 1 cgggatccatgctgggcccc.....tgagctgtctcagaattccg 1017

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_htc:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_gss1:\*  
9: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	921.2	90.6	1635	3 AK002457	AK002457 Mus muscu
2	908.6	89.3	1596	3 AK010857	AK010857 Mus muscu

## ALIGNMENTS

### RESULT 1

AK002457 1635 bp mRNA linear HTC 03-APR-2004  
LOCUS Mus musculus adult male kidney cDNA, RIKEN full-length enriched  
DEFINITION library, clone:0610010D20 product:hypothetical Aminoacyl-transfer  
RNA synthetases class-II/Dihydrodipicolinate synthetase containing  
protein, full insert sequence.  
ACCESSION AK002457  
VERSION AK002457.1 GI:12832454  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE ①  
AUTHORS Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636  
REFERENCE ②  
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20499374  
PUBMED 11042159

FEATURES  
source Location/Qualifiers  
1. .1635  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM DB:0610010D20"  
/db\_xref="taxon:10090"  
/clone="0610010D20"  
/sex="male"  
/tissue\_type="kidney"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
68. .1033  
CDS  
/note="unnamed protein product; hypothetical  
Aminoacyl-transfer RNA synthetases  
class-II/Dihydrodipicolinate synthetase containing protein  
(InterPro|IPR002106, InterPro|IPR002220, evidence:  
InterPro)  
putative"  
/codon\_start=1  
/protein\_id="BAB22114.1"  
/db\_xref="GI:12832455"  
/translation="MLGPQIWASMRQGLSRGLSRNVKGGKVDIAGIYPPVTTTPTATA  
EVDYGLKENLNLRLATFPFRGFVVGSTGEFPFLTSLERLEVSVRQAI PKDKFLIA  
GSGCESTQATVEMTVSMAQVGADVAMVVTVCYRGRMSSAALIHHTYTKVADVSPV  
LYSVPAANTGLELPVDVAVTSLQHPNIIGLKDSGGDVTRIGLIVHKTSKQDFQVLGSA  
GFLASYAVGAVGICGLANVLGAQVCQLERLCLTGQWEAAQELQHRLIEPNTAVTRR  
FGIPGLKKTMDWPGYGGPCRAPLQELSPTEEBALRLDFSNNGWL"

Query Match 90.6%; Score 921.2; DB 3; Length 1635;  
Local Similarity 94.7%; Pred. No. 1.8e-234;

Matches 953; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy	9	ATGCTGGGCCCCAAATCTGGGCCTCCATGAGGCAGGGGCTGAGCAGGGGCTTGTCTAGG	68
Db	68	ATGCTGGGCCCCAAATTTGGGCCTCCATGAGGCAGGGTCTGAGCAGGGGCTTGTCTAGG	127
Qy	69	AACGTGAAGGGAAGAAGATAGACATTGCCGGCATCTACCCACCGTGACCACCCCATTC	128
Db	128	AATGTGAAGGGCAAGAAGGTAGACATTGCCGGCATCTACCCACCGTGACCACCCCATTC	187
Qy	129	ACCGCCACCGCAGAGTAGACTATGGGAACTGGAAGAGAACCTGAACAACTGGCCGCC	188
Db	188	ACCGCCACCGCAGAGGTAGACTATGGGAACTGGAAGAGAACCTGAACAGACTGGCCACC	247
Qy	189	TTCCCTTTTCGAGGCTTCGTGGTCCAGGGCTCTACTGGAGAGTTTCCATTCTTGACCAGC	248
Db	248	TTCCCTTTTCGAGGCTTCGTGGTCCAGGGTTCGACTGGAGAGTTTCCGTTCTTGACCAGC	307
Qy	249	CTTGAGCGCTAGAGGTGGTGGAGCGAGTGCGCCAGGCCATACCAAGGACAAGTCTCTG	308
Db	308	CTCGAGCGCTGGAGGTGGTGGAGCGGTGCGCCAGGCCATACCAAGGACAAGTCTCTG	367
Qy	309	ATAGCCGGCTCTGGCTGCGAGTCCACGCAAGCCACAGTAGAGATGACTGTGACGATGGCT	368
Db	368	ATAGCCGGCTCTGGCTGCGAGTCCACGCAAGCCACAGTAGAGATGACTGTGACGATGGCT	427
Qy	369	CAGGTGGGTGCTGATGTCGCCATGGTGGTGACCCCTTGTTACTATCGCGCCGATGAAC	428
Db	428	CAGGTGGGTGCTGATGTCGCCATGGTGGTGACCCCTTGTTACTATCGTGGCCGATGAGC	487
Qy	429	AGCGCTGCCCTCATTACCACTACACCAAGTTGCTGATCTTTCTCAATCCCGTGGTG	488
Db	488	AGCGCTGCCCTCATTACCACTACACCAAGTTGCTGACGTTTCTCAATCCCTGTTGGTG	547
Qy	489	CTGTACAGTGTCCAGGCAACACGGGTCTAGAGTGCCTGTGGATGCCGTGGTCACATTG	548
Db	548	TTGTACAGTGTCCAGCCAATACGGGCTAGAGTACCTGTGGATGCCGTGGTTACATTG	607
Qy	549	TCTCAGCACCCAAATATCATTGGCTTGAAGGACAGTGGTGGAGATGTGACCAGGACTGGG	608
Db	608	TCTCAGCACCCAAATATCATCGGCTTGAAGGACAGTGGTGGAGATGTGACCAGGATTGGA	667
Qy	609	CTGATTGTTCAAGACAGCAAGCAGGATTTCCAGGTGTTGGCTGGGTGAGTTGGCTTC	668
Db	668	CTGATAGTTCACAGACAGCAAGCAGGATTTCCAGGTGTTGGCTGGGTGAGTTGGCTTC	727
Qy	669	CTCCTGGCCAGCTATGCTGTGGGAGCTGTTGGGGGCATATGTGGCCTGGCCAAATGCTTG	728
Db	728	CTCCTGGCCAGCTATGCTGTGGGAGCTGTTGGGGGCATATGTGGCCTGGCCAAATGCTTG	787
Qy	729	GGGGCCAGGTGTGCCAGCTGGAGAGACTCTGCCTCACAGGGCAGGGGGAAGCTGCCCAG	788
Db	788	GGGGCCAGGTGTGCCAGCTGGAGAGACTCTGCCTCACAGGGCAGTGGGAAGCTGCCCAG	847
Qy	789	AGACTGCAGCACCGCCTCATCGAGCCCAACTGCGGTGACCCGGCGCTTTGGAATACCA	848
Db	848	GAACCTACAGCACCGTCTCATCGAGCCCAACTGCGGTGACCCGGCGCTTTGGAATACCA	907
Qy	849	GGGCTGAAGAAAACCATGGACTGGTTGGGCTACTATGGAGTCCCTGCCGTGCCCCCTTG	908
Db	908	GGGCTGAAGAAAACCATGGACTGGTTGGGCTACTATGGAGTCCCTGCCGTGCCCCCTTG	967
Qy	909	CAGGAGTTGAGCCCTCAGAGGAAGAGGCGCTTCGCTTGGATTTCAGCAACAATGGCTGG	968
Db	968	CAGGAGCTGAGCCCAAGAGGAGGCACTGCGCTTGGATTTCAGCAACAATGGCTGG	1027
Qy	969	CTTTAATGACAAGCGGGGACACCTGGTCTGAGCTGTCTCAGAATT	1014
Db	1028	CTTTAATGACAAGCAGGAGACGCTGGCCTGAGCTATCTGGGACTT	1073

# Result 2

**LOCUS** AK010857 1596 bp mRNA linear HTC 03-APR-2004  
**DEFINITION** Mus musculus 13 days embryo liver cDNA, RIKEN full-length enriched library, clone:2500002N04 product:hypothetical Aminoacyl-transfer RNA synthetases class-II/Dihydrodipicolinate synthetase containing protein, full insert sequence.  
**ACCESSION** AK010857  
**VERSION** AK010857.1 GI:12846588  
**KEYWORDS** HTC; CAP trapper.  
**SOURCE** Mus musculus (house mouse)  
**ORGANISM** Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
**REFERENCE** (1)  
**AUTHORS** Carninci, P. and Hayashizaki, Y.  
**TITLE** High-efficiency full-length cDNA cloning  
**JOURNAL** Meth. Enzymol. 303, 19-44 (1999)  
**MEDLINE** 99279253  
**PUBMED** 10349636  
**REFERENCE** (2)  
**AUTHORS** Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
**TITLE** Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
**JOURNAL** Genome Res. 10 (10), 1617-1630 (2000)  
**MEDLINE** 20499374  
**PUBMED** 11042159

**FEATURES**  
 source Location/Qualifiers  
 1. 1596  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="FANTOM\_DB:2500002N04"  
 /db\_xref="taxon:10090"  
 /clone="2500002N04"  
 /tissue\_type="liver"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="13 days embryo"  
 CDS  
 28. 1020  
 /note="unnamed protein product; hypothetical Aminoacyl-transfer RNA synthetases class-II/Dihydrodipicolinate synthetase containing protein (InterPro|IPR002106, InterPro|IPR002220, evidence: putative)"  
 /codon\_start=1  
 /protein\_id="BAB27226.1"  
 /db\_xref="GI:12846588"  
 /translation="MLGPQIWASMRQGLSRGLSRNVKGMKVDIAGIYPPVTPPTATA  
 EVDYGKLEENLNRLATFPFRGFVVGSTGEFFPLTSLERLEVSRVRQAIPKDKFLIA  
 GSGCESTQATVEMTVSMAQVGADVAVVTPCYRGRMSSAALIHHTYKQADVSPIPVV  
 LYSVPANTGLELPVDAVVTLSQHPNIIIGLKDSSGSDVTRIGLIVHKTQKDFQVLGSA  
 GFLASYAVGAVGGICGLANVLGAQVCQLERLCLTGQWEAAQELQHLIEPKHCGDPA  
 LWNTRAEENHGLVWLLWRSLEPRPVAGAEPHRGGGTALGFPQQWLALMTSRRRLA"

## ORIGIN

Query Match 89.3%; Score 908.6; DB 3; Length 1596;  
 Best Local Similarity 94.5%; Pred. No. 4.2e-231;  
 Matches 952; Conservative 0; Mismatches 54; Indels 1; Gaps 1;

Qy	9	ATGCTGGGCCCCAAATCTGGGCTCCATGAGGCAGGGGCTGAGCAGGGGCTTGTCTAGG	68
Db	28	ATGCTGGGCCCCAAATTTGGGCTCCATGAGGCAGGGTCTGAGCAGGGGCTTGTCTAGG	87
Qy	69	AACGTGAAGGGGAAGAAGATAGACATTGCCGGCATCTACCCACCCGTGACCACCCCATTC	128
Db	88	AATGTGAAGGGCATGAAGGTAGACATTGCCGGCATCTACCCACCCGTGACCACCCCATTC	147
Qy	129	ACCGCCACCGCAGAAGTAGACTATGGGAACTGGAAGAGAACCTGAACAACTGGCCGCC	188
Db	148	ACCGCCACCGCAGAGGTAGACTATGGGAACTGGAAGAGAACCTGAACAGACTGGCCACC	207
Qy	189	TTCCCTTTTCGAGGCTTCGTGGTCCAGGGCTCTACTGGAGAGTTTCCATTCTGACCAGC	248
Db	208	TTCCCTTTTCGAGGCTTCGTGGTCCAGGGTTCGACTGGAGAGTTTCCGTTCCTGACCAGC	267
Qy	249	CTTGAGCGCTAGAGGTGGTGAGCCGAGTGCGCCAGGCCATACCCAGGACAGCTCCTG	308
Db	268	CTCGAGCGCTTGAGGTGGTGAGCCGCGTGCGCCAGGCCATACCCAGGACAGTTCTCTG	327
Qy	309	ATAGCCGGCTCTGGCTGCGAGTCCACGCAAGCCACAGTAGAGATGACTGTCAGCATGGCT	368
Db	328	ATAGCCGGCTCTGGCTGCGAGTCCACGCAAGCCACAGTAGAGATGACTGTCAGCATGGCT	387
Qy	369	CAGGTGGGTGCTGATGCCGCCATGGTGGTGACCCCTTGTTACTATCGCGGCCGATGAAC	428
Db	388	CAGGTGGGTGCTGATGCGCCATGGTGGTGACCCCTTGTTACTATCGTGGCCGATGAGC	447

Qy	429	AGCGCTGCCCTCATTCACTACTACCAAGTTGCTGATCTTTCTCCAATCCCGTGGTG	488
Db	448	AGCGCTGCCCTCATTCACTACTACCAAGTTGCTGACGTTTCTCCAATCCCTGTGGTG	507
Qy	489	CTGTACAGTGTCCAGGCAACACGGGTCTAGAGCTGCCTGTGGATGCCGTGGTCACATTG	548
Db	508	TTGTACAGTGTCCAGCCAATACGGGGCTAGAGCTACCTGTGGATGCCGTGGTTACATTG	567
Qy	549	TCTCAGCACCCAAATATCATTGGCTTGAAGGACAGTGGTGGAGATGTGACCAGGACTGGG	608
Db	568	TCTCAGCACCCAAATATCATCGGCTTGAAGGACAGTGGTGGAGATGTGACCAGGATTGGA	627
Qy	609	CTGATTGTTTCAAGACCAGCAAGCAGGATTTCCAGGTGTTGGCTGGGTGAGTTGGCTTC	668
Db	628	CTGATAGTTTCAAGACCAGCAAGCAGGATTTCCAGGTGTTGGCTGGGTGAGTTGGCTTC	687
Qy	669	CTCCTGGCCAGCTATGCTGTGGGAGCTGTTGGGGGCATATGTGGCCTGGCCAAATGCTTG	728
Db	688	CTCCTGGCCAGCTATGCTGTGGGAGCTGTTGGGGGCATATGTGGCCTGGCCAAATGCTTG	747
Qy	729	GGGGCCCGAGTGTGCCAGCTGGAGAGACTCTGCCTCACAGGGCAGGGGAAGCTGCCCAG	788
Db	748	GGGGCCCGAGTGTGCCAGCTGGAGAGACTCTGCCTCACAGGGCAGTGGGAAGCTGCCCAG	807
Qy	789	AGACTGCAGCACCGCCTCATCGAGCCC-AACACTGCGGTGACCCGCGCTTTGGAATACC	847
Db	808	GAACTACAGCACCGTCTCATCGAGCCCAAACTGCGGTGACCCGCGCTTTGGAATACC	867
Qy	848	AGGGCTGAAGAAAACCATGGACTGGTTTGGCTACTATGGAGGTCCCTGCCGTGCCCTTT	907
Db	868	AGGGCTGAAGAAAACCATGGACTGGTTTGGCTACTATGGAGGTCCCTGCCGTGCCCTTT	927
Qy	908	GCAGGAGTTGAGCCCTCAGAGGAAGAGGCGCTTCGCTTGGATTTAGCAACAATGGCTG	967
Db	928	GCAGGAGTTGAGCCCTCAGAGGAGGAGGCACTGCGCTTGGATTTAGCAACAATGGCTG	987
Qy	968	GCTTTAATGACAAGCGGGGACACCTGGTCTGAGCTGTCTCAGAATT	1014
Db	988	GCTTTAATGACAAGCAGGAGACGCTGGCTGAGCTATCTGGGACTT	1034